STAT 401 Chapter 5.1–5.7

Zepu Zhang
October 12, 2010

A **matrix** is a rectangular array of numbers, arranged in rows and columns. Matrices are used extensively in math and statistics. They provide a convenient way to express linear statistical models, as we shall see in the later sections of chapter 5. The matrix formulation for simple linear regression (linear regression with just one predictor) is identical to the matrix formulation for multiple linear regression (regression using two or more predictors). This makes the transition from simple linear regression (SLR) to multiple linear regression (MLR) smooth and seamless. Formulas for SLR work for MLR as well.

Matrix notation:
(1) conventional bracket notation;
(2) upper-case letter notation; (sometimes boldface; but we don't enforce it)
(3) element listing: $\boldsymbol{A} = [a_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, n$;
(4) dimension (size): $m \times n$, m by n; (5) column vectors and row vectors: often denoted by boldface lower-case letters.

1 Matrices

Here are several matrices that I'll use as examples:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & 7 & -2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

$$E = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}, \quad F = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 3 \\ 3 & 5 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 6 & 10 \end{bmatrix}.$$

The **dimension** or **size** of a matrix is described by the number of rows and the number of columns, separated with the word "by" or the symbol "×". For example, the dimensions of A, B, C, and D are:

$$2 \times 2, 2 \times 3, 1 \times 3, \text{ and } 3 \times 1.$$

The number of rows is always listed before the number of columns.

A **square matrix** is one which has the same number of rows and columns. Examples: A, E and G are square matrices.

A **column vector** is a matrix with just one column, and a **row vector** is a matrix with just one row. Examples: C is a row vector; D and F are column vectors.

The **transpose** of a matrix is a matrix in which the rows and columns are interchanged, and it is denoted by either a 'prime' symbol (as in A') or a superscript T or a superscript t (as in A^T or A^t). Examples:

$$A' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, B' = \begin{bmatrix} 3 & 2 \\ -1 & 0 \\ 2 & 1 \end{bmatrix}, \text{ and } C' = \begin{bmatrix} 4 \\ 7 \\ -2 \end{bmatrix}$$

Two matrices are said to be equal if they have the same dimensions and the entries in one are identical to the entries in the other.

A **symmetric matrix** is a square matrix which is equal to its transpose. Example: G is a symmetric matrix; A and E are not.

2 Matrix addition and subtraction

If two matrices have the same dimensions, it is possible to add (subtract) them; the resulting matrix is the same size, and the entries of the result are obtained by adding (subtracting, respectively) corresponding entries.

For the matrices listed above, A, E and G can be added to or subtracted from one another; B and H can be added to or subtracted from one another. Examples:

$$A+E = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1+2 & 2-1 \\ 3+1 & 4+3 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 4 & 7 \end{bmatrix}$$

$$A-E = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1-2 & 2-(-1) \\ 3-1 & 4-3 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 2 & 1 \end{bmatrix}$$

3 Matrix multiplication

a) Multiplying a matrix by a scalar:

The word 'scalar' is a technical term for a number. When multiplying a matrix by a scalar, simply multiply all the entries in the matrix by the scalar. The scalar is always written to the left of the matrix.

For example,

$$5A = 5\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (5)(1) & (5)(2) \\ (5)(3) & (5)(4) \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix}$$

$$-2B = -2\begin{bmatrix} 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -6 & 2 & -4 \\ -4 & 0 & -2 \end{bmatrix}$$

b) Multiplying a row vector by a column vector (where the row vector is written first and the column vector is written second): This operation is defined when (and only when) the vectors have the same number of entries. If the product is defined, its value is given by summing the products of the entries in the two vectors.

Example: We note that C is a row vector, and D and F are column vectors. The product CD is defined (has meaning) because both C and D have 3 entries; the product CF is not defined. We have:

$$CD = \begin{bmatrix} 4 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = (4)(1) + (7)(2) + (-2)(3) = 12.$$

c) Multiplying one matrix by another: Some products are possible, others are not. Whether a product is defined or not depends on the sizes of the two matrices.

You might guess at this point that matrix multiplication is much like matrix addition and subtraction, and that we could only multiply two matrices if they had the same dimension. For example you might think that

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} * \begin{bmatrix} 2 & -1 & 0 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} (1)(2) & (2)(-1) & (3)(0) \\ (4)(1) & (5)(3) & (6)(5) \end{bmatrix} = \begin{bmatrix} 2 & -2 & 0 \\ 4 & 15 & 30 \end{bmatrix}$$

This type of multiplication exists. This is called the 'Hadamard product.' It is used by a small number of mathematicians. We do not use it.

Instead, we use the standard matrix multiplication, which may initially seem somewhat contrived and artificial; but it actually has some marvellous mathematical properties, as we shall see.

If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then:

1. The product AB is defined if (and only if) $n = p$.
2. If $n = p$, then:
   (a) The product AB is defined and its size is $m \times q$.
   (b) The entry in the r^{th} row and c^{th} column of AB is the product of the r^{th} row of A and the c^{th} column of B (as defined in (b), at the bottom of the previous page).

Note: If $n = p$ so that the product AB is defined, the matrices A and B are said to be 'conformable.'

Here's an easy way to recognize a pair of conformable matrices: Write down their dimensions in a row; this is a list of 4 numbers. If the inner pair of numbers are the same, then the matrices are conformable; and in this event, the outer pair of numbers gives the size of the resulting product matrix.

Example: From before,

$$B = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

B is 2×3 and D is 3×1. So we write:

$$(2 \times 3)(3 \times 1).$$

The inner pair of numbers are the same (both are 3), so the product is defined. And the resulting matrix, BD, is 2×1.

The entry in the (1,1) position (first row, first column) of BD is:

$$\begin{bmatrix} 3 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = (3)(1) + (-1)(2) + (2)(3) = 7.$$

The entry in the (2,1) position (second row, first column) of BD is:

$$\begin{bmatrix} 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = (2)(1) + (0)(2) + (1)(3) = 5.$$

Therefore $BD = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$.

d) Dividing one matrix by another: There is no such thing as matrix division, so this is a very short section.

Usually

$$\boldsymbol{AB} \neq \boldsymbol{BA}.$$

Unless both $\boldsymbol{A}$ and $\boldsymbol{B}$ are square (and of the same size), the two sides are not both defined! When both sides are defined, they usually are still unequal.

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

if the sizes are conformable. Because of this property, we can write $\boldsymbol{ABC}$ and the result is not ambiguous.

$$\boldsymbol{C}(\boldsymbol{A} + \boldsymbol{B}) = \boldsymbol{CA} + \boldsymbol{CB}$$

4 Special types of matrices

Symmetric matrices: A matrix A is **symmetric** if it is equal to its transpose: $A = A^T$. Pictorially, if you draw a line through the matrix from the upper left corner to the lower right corner, the lower portion of the matrix is the mirror image of the upper portion of the matrix.

Example. This matrix is symmetric:

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}$$

• Fact: All symmetric matrices are square (that is, the number of rows is equal to the number of columns).

Diagonal matrices: A matrix is **diagonal** if it is square and all its off-diagonal entries are zero.

Examples. Both of the following matrices are diagonal:

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \text{ and } I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Identity matrix: A diagonal matrix whose main-diagonal entries are all 1 is called an **identity matrix**. We use 'I' to denote an identity matrix. (The same letter is used, whether the matrix is 2×2 or 3×3, etc.)

Fact: If A is an $n \times n$ square matrix, and I is the $n \times n$ identity matrix, then $AI = IA = A$.

The identity matrix I plays the same role for matrices that the number 1 does for real numbers: Multiplying a number by 1 (on the left or right) leaves that number unchanged. Multiplying a square matrix by I (on the left or right) leaves that matrix unchanged.

Vector of 1's: $\mathbf{1}$

Matrix of 1's: $\boldsymbol{J}$ (page 187).
(This does not appear to have a standard notation. Encountered less often.)

All zero's: $\mathbf{0}$.
(In addition and subtraction, $\mathbf{0}$ changes nothing. In multiplication, $\mathbf{0}$ destroys everything; but it may say something about the size of the resultant matrix.)

5 Inverse of a matrix

Given square matrix $\boldsymbol{A}$, if there exists $\boldsymbol{B}$ such that

$$\boldsymbol{AB} = \boldsymbol{I}$$

then $\boldsymbol{A}$ and $\boldsymbol{B}$ are the "inverse (matrix)" of each other. We say "invert" a matrix, or "take the inverse". We use $\boldsymbol{A}^{-1}$ to denote the inverse of $\boldsymbol{A}$.

If someone asks whether two particular square matrices are inverses of one another, the easiest thing to do is to multiply them together (in either order). If the product is the identity matrix, then the original two matrices are inverses of one another; otherwise they are not inverses.

For a 2×2 matrix A, there is a simple formula for the inverse. This formula is valid if the determinant, $\det(\boldsymbol{A})$, is non-zero (otherwise A has no inverse):

$$\text{If } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ then } A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Example If $A = \begin{bmatrix} 2 & 1 \\ 4 & -3 \end{bmatrix}$, then $\det(A) = (2)(\text{-}3)\text{-}(1)(4) = \text{-}10$, so

$$A^{-1} = \frac{1}{-10}\begin{bmatrix} -3 & -1 \\ -4 & 2 \end{bmatrix} = \begin{bmatrix} 3/10 & 1/10 \\ 4/10 & -2/10 \end{bmatrix}.$$

The inverse of a matrix can be used to solve certain equations involving matrices. If we have a matrix equation such as

$$A\mathbf{x} = \mathbf{b},$$

where A is $n \times n$ and $\mathbf{x}$ is a $n \times 1$ column vector of unknowns and $\mathbf{b}$ is a $n \times 1$ column vector, then we can solve for $\mathbf{x}$ by multiplying both sides on the left by A^{-1}, like so:

$$A^{-1}A\mathbf{x} = \mathbf{A^{-1}b}.$$

But $A^{-1}A\mathbf{x} = I\mathbf{x} = \mathbf{x}$, so

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Example Solve the following matrix equation:

$$\begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Solution: Let

$$A = \begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix}.$$

Using the formula for the inverse of a 2×2 matrix,

$$A^{-1} = \frac{1}{(0)(-1) - (3)(1)}\begin{bmatrix} -1 & -3 \\ -1 & 0 \end{bmatrix} = \frac{1}{-3}\begin{bmatrix} -1 & -3 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1/3 & 1 \\ 1/3 & 0 \end{bmatrix}.$$

The matrix equation can then be written as follows:

$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Then

$$A^{-1}A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Therefore

$$I\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Therefore

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1 \\ 1/3 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} (1/3)(1) + (1)(2) \\ (1/3)(1) + (0)(2) \end{bmatrix} = \begin{bmatrix} 7/3 \\ 1/3 \end{bmatrix}.$$

Proposition If $\boldsymbol{A}$ has inverse $\boldsymbol{B}$, then
(1) $\boldsymbol{B}$ is unique.

(2) $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$.
(3) $\boldsymbol{A}^T$ has inverse $\boldsymbol{B}^T$.

To show property (3), one only needs to show $\boldsymbol{A}^T\boldsymbol{B}^T = \boldsymbol{I}$. This is apparent since $\boldsymbol{A}^T\boldsymbol{B}^T = (\boldsymbol{BA})^T = \boldsymbol{I}^T = \boldsymbol{I}$.

Proposition If $\boldsymbol{A}$ and $\boldsymbol{B}$ both have inverses, then $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

To show this, notice

$$(\boldsymbol{AB})(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{A}(\boldsymbol{BB}^{-1})\boldsymbol{A}^{-1} = \boldsymbol{AIA}^{-1} = \boldsymbol{AA}^{-1} = \boldsymbol{I}$$

Note: $(\boldsymbol{A}+\boldsymbol{B})^{-1}$ is not $\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}$. (The sum may not even have an inverse.)

Recap of basic points:

1. We speak of the inverse of square matrices only.
2. If the inverse exists, it is unique.
3. A square matrix may not have an inverse (may not be invertible).

The inverse exists iif (if and only if) $\boldsymbol{A}$ is non-singular, and this is equivalent to saying either (1) the determinant of $\boldsymbol{A}$ is nonzero, or (2) $\boldsymbol{A}$ has full rank (these two conditions are equivalent).

6 Rank and determinant

6.1 Rank

Section 5.5 introduces the notions of 'linear independence' (of the rows or columns of a matrix) and 'rank' (of a matrix) in order to provide a basis for Section 5.6, which discusses the concept of the 'inverse of a (square) matrix.'

Rank is a property of any matrix. But here we worry about square matrix only.

The "rank" of $\boldsymbol{A}$ is the number of linearly independent columns (or rows) of it. (Page 188.)

We say "(or rows)" because examining columns and rows will give you the same number.

If the rank of $\boldsymbol{A}_{n\times n}$ is n, we say it has "full rank" and is "non-singular". In order for a matrix to have an inverse (discussed at some length in Section 5.6), the matrix must be square and it must be "of full rank." It then has a non-zero determinant, and is invertible. If a matrix is not 'of full rank,' then it does not have an inverse.

Definition Linear indepdendence, page 188.

You need to know this definition and use it to determin whether two or three simple vectors are linearly independent.

To show dependence, you only need to come up with one counter example. To show independence, you need to prove it—the coefficients must be all zero in order for the linear combination to be the zero vector.

We care about whether a square matrix if of full rank. Its actual number (if not n) rarely enters calculations. This is different from $\det(\boldsymbol{A})$, which appears in the density function of normal distribution, which is, of course, central to statistics.

6.2 Determinant

Each square matrix has a corresponding, unique number, called its "determinant", $\det(\boldsymbol{A})$, or sometimes $|\boldsymbol{A}|$. (The second notation should be used only when we know this number for the matrix in question is guaranteed to be non-negative. This is an important property and some matrices have this property. A type of such matrices (covariance matrices) are very important in statistics.)

The definition of determinant is essentially how to calculate this number. Following the formula, a square matrix results in a unique number. We will see how to calculate the determinant of a 2×2 matrix; but don't worry about how to calculate this number in general.

Because the number is unique, it's a "property" of a matrix, or a "function" (from a matrix to a real value) of a matrix. In this sense I can call it an "operation" on the matrix.

If $\det(\boldsymbol{A}) \neq 0$, the matrix is "good" (for us): it is non-singular and it is invertible.

If $\det(A) = 0$, we say that A is a 'singular' matrix; in this case, A does not have an inverse. ("A is non-invertible.")

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

$\det(A)$ is defined by:

$$\det(A) = ad - bc.$$

7 Some basic results for matrices

- $A + B = B + A$. (Matrix addition is commutative – order does not matter.) (5.25)
- $(A + B) + C = A + (B + C)$. (Matrix addition is associative.) (5.26)
- $(AB)C = A(BC)$. (Matrix multiplication is associative.) (5.27)
- $C(A + B) = CA + CB$. (The distributive property for matrices.) (5.28)
- $k(A + B) = kA + kB$. (Here, k is a scalar. This law says that we can distribute a scalar.) (5.29)
- $(A')' = A$ (or $(A^T)^T = A$.) (If you take the transpose twice, you get back the original matrix.) (5.30)
- $(A + B)^T = A^T + B^T$. (The transpose of a sum is equal to the sum of the transposes.) (5.31)
- $(AB)^T = B^T A^T$. (The transpose of a product is equal to the product of the transposes – written in the reverse order.) (5.32)
- $(ABC)^T = C^T B^T A^T$ (Says that (5.32) extends to 3 (or more) matrices.) (5.33)
- $(AB)^{-1} = B^{-1}A^{-1}$. (The inverse of a product is equal to the product of the inverses – written in the reverse order. A and B must both be square, and both must have an inverse!) (5.34)
- $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ (Says that 5.34 extends to 3 (or more) matrices.) (5.35)
- $(A^{-1})^{-1} = A$. (If you take the inverse of a matrix two times, you get back the original matrix – provided A has an inverse to begin with.) (5.36)
- $(A^T)^{-1} = (A^{-1})^T$. (Whether you take the inverse first or the transpose first does not matter; you get the same answer both ways. A must be an invertible matrix.) (5.37)

To show the last property, notice

$$\boldsymbol{A}^T\left(\boldsymbol{A}^{-1}\right)^T = \left(\boldsymbol{A}^{-1}\boldsymbol{A}\right)^T = \boldsymbol{I}^T = \boldsymbol{I}$$